Focus Perception and Prominence

Focus Perception and Prominence

ARTICLE · NOVEMBER 2000

Source: CiteSeer

2 AUTHORS:

Maria Wolters
The University of Edinburgh
77 PUBLICATIONS **402** CITATIONS
SEE PROFILE

Petra Wagner
Bielefeld University
131 PUBLICATIONS **430** CITATIONS
SEE PROFILE

Focus perception and prominence

Maria Wolters, Petra Wagner

We discuss if *prominence*, the relative degree of perceptual markedness, provides a basis for signalling wide, narrow and contrastive focus in Concept–to–Speech synthesis. The results indicate that prominence can be used for marking narrow object focus and that higher prominence values signal contrastive focus. Further research is required into signalling narrow subject focus using prominences, into the role of verb prominence, and into durational correlates of focus and prominence.

Liefert *Prominenz*, also der relative Grad perzeptueller Markiertheit, eine Grundlage für die akustische Realisierung von weitem, engem und kontrastivem Fokus in der Concept–to–Speech Synthese? Die Ergebnisse weisen darauf hin, daß enger Objektfokus durch Prominenz markiert werden kann und daß hohe Prominenzwerte kontrastiven Fokus signalisieren. Die Beziehung zwischen Prominenzverteilung und sowohl weitem Fokus als auch der Unterscheidung zwischen Subjekt- und Objektfokus sowie die Rolle der Verbprominenz und der Einfluß der Dauer müssen noch näher untersucht werden.

1 Introduction

In Text-to-Speech synthesis, the input is plain text, which may then be analysed syntactically and morphologically before converting it to speech. In Concept–to–Speech synthesis (CTS), on the contrary, the input text is annotated with semantic and pragmatic information. The system then has to provide acoustic cues to semantic and pragmatic information in the synthesised speech signal. To determine direct acoustic correlates of linguistic concepts on the phonetic and prosodic level is very difficult. Ideally, those cues would be specified at a more abstract level of processing, since it is very difficult to determine direct acoustic correlates of linguistic concepts. Portele and Heuft [12] claim that *prominence* "a quantitative parameter of a syllable or a boundary that describes markedness relative to surrounding syllables and boundaries, respectively" which can take values between 0 and 31 for syllables[5],might provide such an interface between linguistic and acoustic processing. The prominence values can then be transformed into acoustic correlates, which allows them to be used as input to a speech synthesis system. In our system, this transformation is based on a decision tree [7].

We are currently investigating whether it is possible to signal focus scope

and focus placement using prominence. The experiment reported here examines which focus types can be implemented by a straightforward algorithm and which require further work.

1.1 The concept of focus

There are many competing definitions of focus in the literature. Basically, "the term *focus* is used [...] to describe prosodic prominences serving pragmatic and semantic functions" [14, p. 271]. In this experiment, we concentrate on *answer foci*. The answer focus of a declarative utterance can be specified by constructing a question that allows the focus-containing utterance as an answer. Examples:

(1) Q: What happened? A: [The captain noticed the alien]$_F$
(2) Q: Who noticed the alien? A: [The captain]$_F$ noticed the alien.
(3) Q: Who did the captain notice? A: The captain noticed [the alien.]$_F$
(4) Q: Did the captain notice the asteroid? A: The captain noticed [the alien.]$_F$.

These questions express a set of (contextually salient) alternatives to the semantic content asserted in the focused constituent of the answer [14]. Either, these alternatives are a set with one of its elements being the semantic content of the focused constituent (1-3), or the set of alternatives consists of only one element distinct from the semantic content of the focused constituent (4). Following the literature (e.g. [13, 10]), we distinguish *wide focus* (as in ex. 1.A), *narrow focus* (as in ex. 2.A and 3.A) and *contrastive focus* (as in ex. 4.A).

1.2 Perception and Production Experiments on Focus

Although researchers apparently agree that focus is prosodically marked in languages such as English and German, its acoustic realisation and perception is still not clear. Production experiments investigating the acoustic realisation of focus such as [4, 10, 1] found correlates in pitch and duration. Many perception experiments are concerned with focus *placement* (e.g. [2, 9]). In the usual setup, sentences with various pitch accent patterns have to be rated according to their acceptability in various contexts. Portele and Heuft [12] have shown that a prominence–based approach can be used to indicate narrow focus. In their study, five synthesised sentences were presented with the highest prominence value (31) on the accented syllable of the word in focus. The remaining syllables received automatically

generated prominence values. Subjects then had to determine which word was focussed by choosing an appropriate question for each stimulus. Using this method, most foci could be reliably identified. Still, it remains unclear how different prominence values influence focus perception, how wide focus relates to prominence and how narrow focus can be distinguished from contrastive focus.

In our experiment, we investigate if and how the three basic types of focus (wide, narrow, contrastive) can be elicited by varying perceptual prominence. Our study differs from [12] in two important aspects: first, we explicitly investigate different types of foci, and second, the level of prominence is systematically varied. Prevost [13] suggests that contrastive focus might just be a very prominent narrow focus. Therefore we permitted contrastive focus as an additional choice for the subjects, but did not explicitly include it into our conditions for generating the stimuli.

2 Method

2.1 Material

Our basic material is the Bonn Prosodic Database of German (BPDG) [6], which has been labelled with prominences for each syllable. From the BPDG, we chose the two following short SPO sentences read by speaker LF, the female voice of our synthesis system:

(1) Hasen verschwinden im Dickicht. (2) Ein Sofa steht an der Wand.
(1) hares vanish in-the thicket. (2) a sofa stands at the wall.

The standard constituent order ensures that no syntactic topicalisation by fronting can interfere with our listeners' judgements. Examples for fronting: "An der Wand steht das Sofa.", "Im Dickicht verschwinden Hasen".

The prominences as labelled in the database (phonemic transcriptions in a slightly modified version of SAMPA[1]) are:

(1) ?aIn zo: fa Ste:t ?an de:6 vant. (2) ha: z@n fE6 SvIn d@n Im dIk ICt
(1) 7 23 5 16 5 4 22 (2) 25 4 3 21 1 5 25 2

These prominences were then varied according to the conditions summarised in Tab. 1. The conditions are motivated by the following very simple focus assignment algorithm for SPO sentences:
Given a set of lexical prominences,

[1] Speech Assessment Methods Phonetic Alphabet,
http://www.phon.ucl.ac.uk/home/sampa/home.htm

	subject prom.	object prom.		subject prom.	object prom.
W	varied	= subj.prom.			
NSA	varied	15	NOA	15	varied
NSN	varied	14	NON	14	varied

Table 1: Conditions for generating the stimuli: subject and object prominence. 15 is the lowest prominence value that still triggers a pitch accent. Since the indication of focus type in our stimuli is rather crude, the terms "wide focus" and "narrow focus" will almost always be preceded by the adjective "intended" when describing classes of stimuli.

1. if no special focus: do not modify subject and object prominence
2. if narrow focus on subject or object: set prominence=V

The purpose of our experiment was to determine which extensions to this algorithm are needed.

The conditions were designed to cover all three types of focus discussed in section 1.1. With condition W, we tested how a baseline setting with subject and object equally prominent would be interpreted. We conjectured that if there was a prominence contour evoking wide focus, it would have to be a comparatively flat one. Therefore, a "wide focus" question was included in the list of potential contexts.

The comparatively small degree of deaccentuation is due to the fact that we wanted to modify the original prominence contour only where necessary in order to preserve a little naturalness. The prominence of the verb was not manipulated in order to restrict the number of parameters. Its influence on focus perception will be the subject of a future experiment.

Each condition was tested with five levels of prominence from 15 (weakly accented) to 31 (fully accented). Prominence level 1 corresponds to a prominence of 15, level 2 to 19, level 3 to 23, level 4 to 27 and level 5 to 31. For level 1 (and level 2, sentence 1), the verb was the most prominent element of the sentence, for levels 3-5, the prominence of the accented word always exceeded that of the verb.

This setup yields 2 (sentences) $\times$ 5 (conditions) $\times$ 5 (levels) = 50 different prominence annotations. Adding the original prominence contours of both sentences yields a total of 52 different stimuli. Each stimulus was presented twice, resulting in 104 stimuli in all.

The stimuli were generated using the Verbmobil speech synthesis system, sampling frequency 32 kHz. The Verbmobil system is a PSOLA–based [8] concatenative synthesis system; for a description of the inventory structure, see [11]. The input was a phonemic transcription of the database in our version of SAMPA, with a prominence value between 0 and 31 associated

Question		evoked
Sentence 1	Sentence 2	focus type
Was sieht man auf dem Bild?		wide (W)
What do you see on the picture?		
Was verschwindet im Dickicht?	Was steht an der Wand?	subject
What's vanishing in the thicket?	*What is there at the wall?*	narrow (NS)
Verschwinden Kaninchen im Dickicht?	Steht ein Bett an der Wand?	subject
Do rabbits vanish in the thicket?	*Is there a bed at the wall?*	contrastive (CS)
Wo verschwinden Hasen?	Wo steht ein Sofa?	object
Where do hares vanish?	*Where is a sofa?*	narrow (NO)
Verschwinden Hasen im Unterholz?	Steht ein Sofa am Fenster?	object
Do hares vanish in the undergrowth?	*Is a sofa at the window?*	contrastive (CO)

Table 2: Questions for evoking different types of foci

with each syllable.

2.2 Design

Five phoneticians including the authors participated in the experiment. Because of the exploratory nature of this experiment, we decided not to use paid naive subjects. The stimuli were presented in two groups, first, those derived from sentence 1, then, those based on sentence 2. The order of the 52 stimuli in each group was randomised. It was the same for all subjects. The subjects listened to the sentences in a quiet room via loudspeakers. They were allowed to repeat to each stimulus as often as they wanted. Subjects were asked: "If the stimulus were the answer to a question, which of the following five questions would that be?" This task definition is based on the operationalisation in section 1.1. Table 2 presents an overview of the questions and the corresponding types of answer foci.

3 Results

As to be expected, we find a highly significant correlation between correct scope and correct placement identification ($0.599, p < 0.0001$). The correlations between correct scope and prominence (0.282) and correct placement and prominence (0.138) are highly significant ($p < 0.001$) but very small. Narrow focus (40.35% correct scope and placement identification) is far easier to identify than wide focus (29% correct scope identification). The original sentences as taken from the database were mostly perceived as having a narrow focus (90%) on the object (55%), but this focus was only

condition	evoked focus type				
	wide	narrow subj.	contr. subj.	narrow obj.	contr. obj.
W	29	12	10	33	16
NSA	37	6	8	26	23
NSN	32	22	1	29	16
NOA	27	6	1	37	29
NON	33	7	1	33	26

Table 3: Judgements in % for each condition. For a list of evoked focus types, refer to Tab. 2. Key: contr.: contrastive, subj.: subject, obj.: object

judged to be contrastive in 20% of all cases.

The results of the subjects sometimes differed quite markedly. One subject perceived disproportionally many wide foci (54,1%) and disproportionally few object foci, while another subject perceived no subject foci. All subjects varied widely in their perception of contrastive foci (41% – 13.3%). Table 3 summarises the subjects' judgements for all conditions.

3.1 Focus scope

The subjects' perception of intended focus scope is summarised in Tab. 4. While intended narrow focus is recognised quite well, condition W is perceived as narrow focus in 71% of all cases. Our hypothesis that a flat contour might be likely to signal wide focus is therefore not corroborated. The overall recognition rate for focus scope is 57,9%, which is due to the high proportion of intended narrow foci in the stimuli. Whenever scope is recognised correctly, the prominence level is significantly higher (two-tailed t-test, $p < 0.001$). For correct scope, the median is at 3, while for incorrect scope, it is at 2. This holds for condition W as well as for the narrow focus conditions NSN, NSA, NOA and NOP (two-tailed t-tests, in both cases $p < 0,0001$).

Does a low prominence level correspond to wide, a high level to narrow focus? We separate the stimuli into two classes, one with prominence level ≥ 3 (Pr3), one with prominence level ≤ 2 (Pr2). For Pr3, there is a clear preference for perceiving narrow focus while for Pr2, results are completely random, with each focus type classified about equally frequently as narrow or wide.

3.2 Focus placement

Subjects' judgements for intended object and subject foci are summarised in Tab. 5. While the stimuli contained 38,5% subject and 38,5% object

intended scope	**perceived scope**					
	All		Pr3		Pr2	
	narrow	wide	narrow	wide	narrow	wide
narrow	67.8	71	80.8	85	48.1	50
wide	32.3	29	19.2	15	51.9	50

Table 4: Judgements in % on intended narrow/wide focus. Key: All: all prominence levels, Pr2: prominence levels ≤ 2, Pr3: prominence levels ≥ 3

intended place	**perceived place**								
	All			Pr3			Pr2		
	W	O	S	W	O	S	W	O	S
object	30	62.5	7.5	15.8	79.2	5	51.3	37.5	17.5
subject	34.5	47	18.5	22.5	58.3	19.2	52.5	30	11.3

Table 5: Judgements on intended subject/object focus. Key: S: subject focus, O: object focus, W: wide focus; All: all prominence levels, Pr2: prominence levels ≤ 2, Pr3: prominence levels ≥ 3

foci, listeners clearly favoured object (53,9%) over subject foci (15,6%). Focus placement is detected correctly for 40.5% of all intended narrow foci. This performance below chance level is mainly due to the bad recognition of intended subject focus (19,6%); the rate for intended object focus is much higher (62.5%). When listeners did not perceive an object focus, the intended focus was frequently a subject or wide focus (69.01%). The pitch accent on the unfocused constituent in conditions NSA, NOA is far too weak to influence the recognition of focus placement (two-tailed Fisher's F, $p < 0.839$).

The distribution of prominences shows a similar pattern as for focus scope, which is, however, less significant. The median prominence level of correctly detected stimuli is 3, whereas for the incorrectly detected ones, it is 2 (two-tailed t-test, $p < 0.03$).

There is an interesting correlation between perceived focus type and prominence level. At higher levels (group Pr3), listeners tend to select object focus, while for lower prominences (group Pr2), subject and wide foci are clearly recognised as non-object foci and object foci are frequently misclassified. This is due to the fact that subjects choose wide focus disproportionally often (28% intended vs. 47.51% perceived wide foci). There are two possible reasons. First, for sentence 1 and levels 1 and 2 and for sentence 1 and level 1, it is the verb which bears the highest pitch accent and not one of its arguments. Secondly, using low prominence levels yields a relatively flat contour. The relevance of each of these factors will be examined in subsequent experiments..

3.3 Contrastive focus

In 26,1% of all cases, the subjects detected contrastive foci. 21,9% of all stimuli perceived as contrastive belonged to condition W. The median prominence level for contrastive focus was 4, which results in a very pronounced accent, as opposed to 2 for non-contrastive foci (two-tailed t-test, $p < 0.0001$). This indicates that the subjects use degree of accentuation to detect contrastivity.

Since high prominences correlate significantly with the detection of object focus, focus placement recognition should be significantly better for contrastive foci than for non-contrastive ones. A two-tailed t-test shows that this prediction is borne out by the data (two-tailed t-test, $p < 0.003$). Contrast focus is placed correctly in 47,56% of all cases, non-contrast focus only in 33,25%. Why is the subjects' performance still worse than chance? High prominences only aid in detecting object, not in detecting subject focus. 83,21% of all contrastive foci were perceived as object foci, but only 41,6% were intended as such. Furthermore, almost all object foci are recognised correctly (96,5%), but that wide and subject foci are quite often classified as object foci (74,38%). It follows that if we want to signal subject focus by a very prominent pitch accent on the subject only, this can be misinterpreted quite easily as object focus. On the other hand, if a subject focus is in fact perceived, the judgement is correct in 91% of all cases.

4 Discussion

The results confirm the conclusion of [12] that using prominence to signal focus scope and focus placement is feasible. A detailed analysis showed that narrow object foci can be indicated quite well by prominence level alone. The higher the prominence, the better narrow object foci are recognised, while very high prominences tend to indicate contrastiveness.

However, we did not find any cues to wide focus. Subjects may have heard disproportionally many wide foci at low prominence levels because for these conditions, the verb tended to bear the main pitch accent. Oppenrieder [10] suggests that wide focus is usually not signalled consistently by speakers. If this is true, further research should concentrate on explicit cues to narrow foci on different constituents.

Contrary to our results for object focus, high prominence values on the subject rarely trigger perception of a subject focus. One subject com-

mented that sometimes she heard a weak subject focus, but considered it too weak to qualify for a narrow focus. Interestingly enough, the prominence levels of subject and object were roughly equal for both original sentences, and this was perceived by most subjects as narrow focus on the object. This indicates that other parts of the prominence contour have to be varied more than we have done here. For example, we might want to deaccent *all* other accented syllables more fully (cf. [4]).

Furthermore, there were no explicit duration cues for subject focus. [3] suggests that increased duration is important for utterance initial focus (here: subject focus), but not for final focus because of final lengthening. This indicates that the relation between prominence values and duration needs further attention.

Finally, the subjects might have been confused because the subject of sentence (1) was a bare plural ("Hasen") and of sentence (2), an indefinite noun phrase. This resulted in somewhat awkward questions. Sentences for further experiments will have to be designed especially for that purpose. We conjecture that the more definite a subject, the easier it is to signal subject focus. If this is true, our simple algorithm was subjected to a worst case test here.

5 Conclusion

The prominence–based approach to speech synthesis allows a rather straightforward modelling of linguistic concepts by degrees of prominence [12, 7], but many perception experiments are needed both to improve the acoustic realization of prominence values and to understand the actual relationship between prominences and linguistic concepts. The contribution of duration will also have to be investigated more thoroughly.

Acknowledgements

We would like to thank our subjects for their help and Thomas Portele, Anja Elsner, Bernhard Schröder, and two anonymous reviewers for their comments. This research was partly funded by the German Federal Ministry of Education, Science, Research, and Technology (BMBF), Verbmobil project, Grant 01 IV 101 G.

References

[1] K. Alter, J. Matiasek, and G Niklfeld. Modeling prosody in a German concept-to-speech system. In D. Gibbon, editor, *Natural Language Processing and Speech Technology. Proc. 3rd KONVENS*, chapter 16, pages 156–165. Mouton de Gruyter, Berlin, 1996.

[2] S. Birch and C. Clifton. Focus, accent, and argument structure: effects on language comprehension. *Language and Speech*, 38(4):365–391, 1995.

[3] W.E. Cooper, S.J. Eady, and P.R. Mueller. Acoustical aspects of contrastive stress in question-answer contexts. *J. Acoust. Soc. Amer.*, 77(6):2142–2156, 1985.

[4] S.J. Eady and W.E. Cooper. Speech intonation and focus location in matched statements and questions. *J. Acoust. Soc. Amer.*, 80(2):402–415, 1986.

[5] G. Fant and A. Kruckenberg. Preliminaries to the study of swedish prose reading and reading style. *STL-QPSR*, 2/3:1–53, 1989.

[6] B. Heuft, T. Portele, F. Höfer, J. Krämer, H. Meyer, M. Rauth, and G. Sonntag. Parametric description of F_0-contours in a prosodic database. In *Int. Cong. Phon. Sci.*, volume 2, pages 378–381. 1995.

[7] Barbara Heuft. *Eine prominenzbasierte Methode zur Prosodieanalyse und -synthese.* PhD thesis, Institut für Kommunikationsforschung und Phonetik der Universität Bonn, 1998.

[8] E. Moulines and F. Charpentier. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 9:453–467, 1990.

[9] S.G. Nooteboom and J.G. Kruyt. Accents, focus distribution, and the perceived distribution of given and new information: an experiment. *J. Acoust. Soc. Amer.*, 82:1512–1524, 1987.

[10] W. Oppenrieder. Fokus, Fokusprojektion und ihre intonatorische Kennzeichnung. In H. Altmann, A. Batliner, and W. Oppenrieder, editors, *Zur Intonation von Modus und Fokus im Deutschen*, pages 267–280. Niemeyer, Tübingen, 1989.

[11] T. Portele. *Ein phonetisch-akustisch motiviertes Inventar zur Sprachsynthese deutscher Äußerungen.* Niemeyer, Tübingen, 1996.

[12] T. Portele and B. Heuft. Towards a prominence-based speech synthesis system. *Speech Communication*, 21:61–72, 1997.

[13] S. Prevost. *A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation.* PhD thesis, University of Philadelphia, 1995.

[14] Mats Rooth. Focus. In S. Lappin, editor, *Handbook of Contemporary Semantic Theory*, pages 271–297. Blackwell, Oxford, 1995.